# NATURAL LANGUAGE PROCESSING

ada kiligilik kelabatan jajata. Lib	PHOCESSING
-> Roadmap . mileson	state to assessment to all to assess
Machine Learning	Supervised Learning -> Regression -> Supervised Learning -> Classification -> Unsupervised Learning
$f_1$ $f_2$ $f_3$	fy output > Model
	Prediction
marker dans brows	Text?
Spam Classificat	tion -> Email - > Not Spam
Email Subject E	Email Body Spam/Ham
Tanisman Mossiles &	morforegani

Vectors (Convert) Convert to modern comment of half &

assimme di

(institution)

		e in it,	1000	Î			
Cleaning T	Text Pr	eprocessing	)→Tok	eniza	tion, Lemn	natization, Stemm	nin
s Souperstor 4, 11 month action		l labinaga	<u>J.</u>				
Vectors	Text Pre	processing -	-> BOW,	TF-10	of, Unigra	m, Bigram,	
•	Maran .	i ser raguen'	1				
Vectors	Text Prep	rocessing —	-> Worda	evec,	CBOW, Shi	p-gram,	
(New Yorks)		deglas :	V Mi		C	\$	
Deep Learning	BNNs, LS	STM, GRU,		_)			
			Ţ				
	Tex	+ Preprocess	sing —> L	brow	Embedding	13	
Silve			,			<del>-,</del> .	
	mrc <sup>5</sup> d	Transf	ormers				
1	moge long.	BER-	, (rel-	nos 🎚	Bendi) y	nog	
_> Use C	ases	υ'.	3 iran	7	i i		
D Spelling	correction			6		Answering	
3 Translat				7	Google	Assistant	
3 Text to	images /v	beos	Vivini			J.	
9 Summari		212				(America)	
6 Classifi	cation ( Tex	4)					

DESCRIPTION LANGUAGE

#### -> Tokenization

- 1 Corpus -> Paragraph
- @ Documents -> Sentences
- 3 Vocabulary -> Unique words
- 4) words -> words

Tokenization is a process
of converting a
paragraph into sentences
and sentences into words

Corpus { "My name is Saad and I have a interest in Machine Learning. I am also a Dade Science Consultant."

## Tokens Esentences }

- D My name is saad and I have a interest in Machine Learning
- @ 1 am also à Data Science Consultant

#### N

#### Tokenization {words}

- 1) My, name, 15, Saad, and, I, have, a, interest, in, Machine, Learning
- 1, am, also, a, Data, Science, Consultant

### -> Stemming and Lemmatization

## > Stemming

It is the process of removing the last few characters of a given word, to obtain a shorter form, even if that form doesn't have any meaning

Markeya & Land Control of

History Final Final

## > Lemmatization

It is a linguistic process that involves.

reducing words to their base or roof form

Known as lamma.

History Final History